526 Rec'd PCT/PTO 12 JAN 2001

SUBSTITUTE FORM PTO-1390 U S DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

# TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

ATTORNEY'S DOCKET NUMBER
08106-004001

U.S. APPLICATION NO. (If Known, see 37 CFR 1.5) 09/743731

| INTERNATIONAL APPLICATION NO. | INTERNATIONAL FILING DATE | PRIORITY DATE CLAIMED |
|---|---|---|
| PCT/CA99/00637 | 14 July 1999 | 14 July 1998 |

TITLE OF INVENTION
CLEAVAGE OF CAULOBACTER PRODUCED RECOMBINANT FUSION PROTEINS

APPLICANT(S) FOR DO/EO/US
John Smit

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☒ This is an express request to promptly begin national examination procedures (35 U.S.C. 371(f)).
4. ☒ The US has been elected by the expiration of 19 months from the priority date (PCT Article 31).
5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. ☒ is attached hereto (required only if not communicated by the International Bureau).
   b. ☐ has been communicated by the International Bureau.
   c. ☐ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☐ An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).
7. ☐ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. ☐ are attached hereto (required only if not communicated by the International Bureau).
   b. ☐ have been communicated by the International Bureau.
   c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
   d. ☐ have not been made and will not be made.
8. ☐ An English language translation of amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9. ☐ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. ☐ An English language translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

**Items 11 to 16 below concern other documents or information included:**

11. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. ☐ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. ☒ A FIRST preliminary amendment.
    ☐ A SECOND or SUBSEQUENT preliminary amendment.
14. ☐ A substitute specification.
15. ☐ A change of power of attorney and/or address letter.
16. ☐ Other items or information:
    ☐
    ☐
    ☐
    ☐
    ☐

**CERTIFICATE OF MAILING BY EXPRESS MAIL** Express Mail Label No EL270011425US

I hereby certify under 37 CFR §1 10 that this correspondence is being deposited with the United States Postal Service as Express Mail Post Office to Addressee with sufficient postage on the date indicated below and is addressed to the Commissioner for Patents, Washington, D C 20231.

| January 12, 2001 | Samantha Bell | Samantha Bell |
|---|---|---|
| Date of Deposit | Signature | Typed Name of Person Signing |

JC07 Rec'd PCT/PTO 12 JAN 2001

| U.S. APPLICATION NO. (IF KNOWN) 09/743731 | INTERNATIONAL APPLICATION NO. PCT/CA99/00637 | ATTORNEY'S DOCKET NUMBER 08106-004001 |
|---|---|---|

17. ☐ The following fees are submitted:

**Basic National Fee ( 37 CFR 1.492(a)(1)- (5) ):**

Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO ............ **$1000**

International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO ........ **$860**

International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO ..................... **$710**

International preliminary examination fee paid to USPTO (37 CFR 1.482) but all claims did not satisfy provisions of PCT Article 33(1)-(4) ................. **$690**

International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(1)-(4) ........................... **$100**

| | | | | CALCULATIONS | PTO USE ONLY |
|---|---|---|---|---|---|
| **ENTER APPROPRIATE BASIC FEE AMOUNT =** | | | | $860.00 | |
| Surcharge of **$130** for furnishing the oath or declaration later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(e)). | | | | $0.00 | |
| Claims | Number Filed | Number Extra | Rate | | |
| Total Claims | 8 - 20 = | 0 | x **$18** | $0.00 | |
| Independent Claims | 2 - 3 = | 0 | x **$80** | $0.00 | |
| MULTIPLE DEPENDENT CLAIMS(S) (if applicable) | | | + **$270** | $0.00 | |
| **TOTAL OF ABOVE CALCULATIONS =** | | | | $860.00 | |
| ☐ Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2. | | | | $0.00 | |
| **SUBTOTAL =** | | | | $860.00 | |
| Processing fee of **$130** for furnishing the English Translation later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(f)) | | | | $0.00 | |
| **TOTAL NATIONAL FEE =** | | | | $860.00 | |
| Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31).**$40.00** per property + | | | | $0.00 | |
| **TOTAL FEES ENCLOSED =** | | | | $860.00 | |
| | | | | **Amount to be refunded:** | $ |
| | | | | **Charged:** | $ |

a. ☒ A check in the amount of $860.00 to cover the above fees is enclosed.

b. ☐ Please charge my Deposit Account No. 06-1050 in the amount of $0.00 to cover the above fees. A duplicate copy of this sheet is enclosed.

c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 06-1050. A duplicate copy of this sheet is enclosed.

**NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b) must be filed and granted to restore the application to pending status.**

SEND ALL CORRESPONDENCE TO:

John T. Li
FISH & RICHARDSON P.C.
225 Franklin Street
Boston, MA 02110-2804
(617) 542-5070 *phone*
(617) 542-8906 *facsimile*

John T. Li
SIGNATURE:

John T. Li
NAME

44,210
REGISTRATION NUMBER

09/743731

Attorney's Docket No.: 08106-004001 / 82104-17

JC07 Rec'd PCT/PTO 12 JAN 2001

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

| | | | |
|---|---|---|---|
| Applicant | : John Smit | Art Unit | : Unknown |
| Serial No. | : | Examiner | : Unknown |
| Filed | : Herewith | | |
| Title | : CLEAVAGE OF CAULOBACTER PRODUCED RECOMBINANT FUSION PROTEINS | | |

**Box PCT**
Commissioner for Patents
Washington, D.C. 20231

## PRELIMINARY AMENDMENT

Prior to examination, please amend the application as follows:

In the Claims:

In claim 3, line 15, delete "or 2".

In claim 4, line 18, delete "or 2".

In claim 5, line 21, please delete "any one of claims 1-4" and insert therefore --claim 1--.

In claim 6, line 24, please delete "any one of claims 1-5" and insert therefore --claim 1--.

In claim 7, lines 27-28, please delete "suitable for use in the method of claim 1, wherein the method".

In claim 8, line 8, please delete "as described in" and insert therefore --by the method of--.

CERTIFICATE OF MAILING BY EXPRESS MAIL

Express Mail Label No. EL270011425US

I hereby certify under 37 CFR §1.10 that this correspondence is being deposited with the United States Postal Service as Express Mail Post Office to Addressee with sufficient postage on the date indicated below and is addressed to the Commissioner of Patents, Washington, D.C. 20231.

January 12, 2001
Date of Deposit

Samantha Bell
Signature

Samantha Bell
Typed or Printed Name of Person Signing Certificate

REMARKS

All amendments are to remove multiple dependencies or to clarify the claims language. No new matter has been added.

Applicant submits that all of the claims are now in condition for examination, which action is requested. Please apply any charges or credits to Deposit Account No. 06-1050.

Respectfully submitted,

Date: 1-12-01

John T. Li

John T. Li
Reg. No. 44,210

Fish & Richardson P.C.
225 Franklin Street
Boston, MA 02110-2804
Telephone: (617) 542-5070
Facsimile: (617) 542-8906

20184130.doc

09/743731



JC07 Rec'd PCT/PTO 12 JAN 2001

# CLEAVAGE OF CAULOBACTER PRODUCED RECOMBINANT FUSION PROTEINS

## FIELD OF INVENTION

This invention relates to the expression and secretion of recombinant fusion proteins from Caulobacter wherein a heterologous polypeptide is fused with all or part of the surface layer protein (S-layer protein) of the bacterium.

## BACKGROUND OF THE INVENTION

Many bacteria assemble layers composed of repetitive, regularly aligned, proteinaceous sub-units on the outer surface of the cell. These layers are essentially two-dimensional paracrystalline arrays, and being the outer molecular layer of the organism, directly interface with the environment. In Caulobacter, the S-layer protein is synthesized by the cell in large quantities and the S-layer completely envelops the cell and thus appears to be a protective layer.

Caulobacter are natural inhabitants of most soil and freshwater environments and may persist in waste water treatment systems and effluents. The bacteria alternate between a stalked cell that is attached to a surface, and an adhesive motile dispersal cell that searches to find a new surface upon which to stick and convert to a stalked cell. The bacteria attach tenaciously to nearly all surfaces and do so without producing the extracellular enzymes or polysaccharide "slimes" that are characteristic of most other surface attached bacteria. Caulobacters have simple requirements for growth. The organism is ubiquitous in the environment and has been isolated from oligotrophic to mesotrophic situations. They are known for their ability to tolerate low nutrient level stresses, for example, low phosphate levels.

All of the freshwater Caulobacter that produce an S-layer are similar and have S-layers that are substantially the same under election microscopy. The layers are hexagonally arranged in all cases, with a similar centre-centre dimension (see: Walker, S.G., et al.. (1992). "Isolation and Comparison of the Paracrystalline Surface Layer Proteins of Freshwater Caulobacters" J. Bacteriol. 174: 1783-1792).

16S rRNA sequence analysis of several S-layer producing Caulobacter strains show that they group closely (see: Stahl, D.A. et al. (1992) "The Phylogeny of Marine and Freshwater Caulobacters Reflects Their Habitat" J. Bacteriol. 174: 2193-2198). DNA probing of Southern blots using the S-layer gene from C. crescentus CB15 identifies a single band that is consistent with the presence of a cognate gene (see: MacRae, J.D. and. J. Smit. (1991) "Characterization of Caulobacters Isolated from Wastewater Treatment Systems" Applied and Environmental Microbiology 57:751-758). Furthermore, antisera raised against the S-layer protein of CB15 reacts against the S-layer protein of other Caulobacter (see: Walker, S.G. et al. (1992) [*supra*]). All S-layer proteins isolated from Caulobacter may be substantially purified using the same methods. All strains appear to have a polysaccharide species which may be required for S-layer attachment (see: Walker, S.G. et al. (1992) [*supra*]).

The S-layers elaborated by freshwater isolates of Caulobacter are visibly indistinguishable from the S-layer produced by Caulobacter strains CB2 and CB15. The S-layer proteins from the latter strains have approximately 100,000 m.w. although sizes of S-layer proteins from other species and strains will vary. The hydrophillic S-layer protein has been characterized both structurally and chemically. It is composed of ring-like structures spaced at 22 nm intervals arranged in a hexagonal manner on the outer membrane The S-layer is bound to the bacterial surface and may be removed by low pH treatment or by treatment with a calcium chelator such as EDTA.

The similarity of S-layer proteins in different strains of Caulobacter permits the use of a cloned S-layer protein gene of one Caulobacter strain for retrieval of the corresponding gene in other Caulobacter strains (see: Walker, S.G. et al. (1992) [*supra*]; and MacRae, J.D. et al. (1991) [*supra*]).

Expression of a heterologous polypeptide as a fusion product with the S-layer protein of Caulobacter provides advantages not previously seen in systems for production of recombinant fusion proteins using other organisms such as E. coli and Salmonella. All known Caulobacter strains are believed to be harmless and are nearly ubiquitous in aquatic environments. In contrast, many Salmonella and E. coli strains are pathogens. Consequently, expression and secretion of a heterologous polypeptide using Caulobacter as a vehicle has the advantage that the expression system will be

stable in a variety of outdoor environments and may not present problems associated with the use of a pathogenic organism. Furthermore, Caulobacter are natural biofilm forming species and may be adapted for use in fixed biofilm bioreators. The quantity of S-layer protein that is synthesized and is secreted by Caulobacter is high, reaching 12% of the cell protein.

There is an existing need to produce pure proteins and peptides in an economical manner and in a manner that minimizes or simplifies the purification steps needed after fermentation. Key commercial areas include the production of recombinant human and animal therapeutic antibiotic and vaccine peptides, industrial enzymes, protein polymers, and antibacterial enzymes for foodstuffs. Many of these commercial applications require low production costs and there are few expression systems available that can meet such cost restraints. In addition, there are numerous research applications where rapid methods to produce and purify proteins are needed to facilitate the discovery stage. This is especially true where there is a desire to express a large number of proteins with unknown function (from a collections of cloned cDNA's, for example) or a large number of variants of a single protein, (for example, resulting from site directed mutagenesis) in a search for variants with improved properties.

Generally, proteins must be secreted to be produced at low cost. The primary reason is the much reduced cost of purification of the target protein from cell material. However, even for secreted proteins, simple methods of separating the product from spent culture and cells are important for cost reduction and ease of use.

An international patent application published as WO 97/34000 on September 18, 1997 describes the expression and secretion of recombinant proteins from Caulobacter in which the recombinant protein is a fusion of all or part of Caulobacter S-layer protein with a heterologous protein of interest (also see: Bingle, W.H., et al. 1997[1] "Linker Mutagenesis of the Caulobacter us S-layer protein: Toward a Definition of an N-terminal Anchoring Region and a C-terminal Secretion Signal and the Potential for Heterologous Protein Secretion". J. Bacteriol. 179:601-611).

The Caulobacter S-layer secretion apparatus is in the category of "Type 1" secretion usually found in pathogenic bacteria and noted for its ability to secrete a wide variety of proteins including large and hydrophillic proteins. The Caulobacter protein

secretion system is particularly useful to secrete recombinant proteins.

The Caulobacter S-layer Type 1 secretion pathway requires only a C-terminal secretion signal, typically comprising about 200 amino acids at the end of the protein. The export mechanism is capable of tolerating a wide variety of foreign proteins. Recombinant proteins may be conveniently produced as fusion proteins with the target protein being fused to the C-terminal secretion signal. Depending on the application, it may be desirable to remove the secretion signal following secretion. Not removing the secretion signal may be an approach suitable for many subunit vaccine applications, where the remaining S-layer protein serves as a carrier.

A unique and desirable feature of fusion proteins produced by the Caulobacter S-layer protein secretion system is that they form insoluble aggregates in the culture medium. This is apparently a consequence of the S-layer sequences associated with secretion signal and reflects the fact that the protein normally self-assembles into a two dimensional crystalline layer on the bacterium's surface. These aggregates are visible to the naked eye and are readily collected by simple filtration. With simple water wash steps, residual bacterial cells are readily flushed away. It is routinely possible to achieve a protein purity of 90% or better with this simple purification procedure.

## DESCRIPTION OF THE PRIOR ART

Most current protein purification systems for recombinant proteins produced by bacteria rely upon an affinity matrix to achieve separation of the target protein and to concentrate the protein for subsequent steps of purification. To accomplish this, genes for recombinant proteins are commonly constructed so that they contain affinity tags, which are protein sequences that will bind to an affinity matrix. Commonly used systems include the following:

(a) glutathione S-transferase (GST) tag, which binds to glutathione-sepharose matrices;

(b) maltose binding protein (MBP) tag, which binds to amylose matrices;

(c) multiple tandem histidine residues (e.g. "His-6") tag, which binds to nickel-derivatized solid matrices; and

(d) protein A tag, which binds to Immunoglobulin IgG-derivatized sepharose or comparable matrices.

Prior art techniques were typically developed so that removal of a target protein does not disrupt the tag and matrix association. Instead, enzymes that cleave specific sequences of amino acids are employed. The enzyme cleavage sequence is positioned between the tag and the desired recombinant protein and enzymatic cleavage is effected directly on the matrix with attached fusion protein. If a secretion signal is used, the cleavage site is usually positioned such that the secretion signal is separated from the target recombinant protein during the cleavage step. The matrix is regenerated for re-use only after the target recombinant protein has been purified away from the matrix. Typical enzymes used in these methods are Factor Xa, enterokinase and collagenase.

Chemical cleavage is generally not used because the conditions required for cleavage will disrupt the binding of affinity tag and matrix or destroy the matrix. When chemical cleavage is used with recombinant fusion proteins to cleave target protein from a secretion signal and/or affinity tag, solubilization and denaturation processes are generally employed. The expectation is that complete or nearly complete unfolding of the protein is a prerequisite for effective cleavage.

Mild-acid cleavage is predicated on the inclusion, by happenstance or design, of the acid-sensitive aspartate-proline dipeptide at a desired site for cleavage. The protein to be cleaved is typically exposed to conditions that solubilize and/or completely denature the protein prior to cleavage. The chaotropic agent guanidine hydrochloride (used at 6-7 M) is commonly employed to denature and solubilize the protein prior to, or at the same time as acid treatment. Alternately, high concentrations of acids that also serve as solubilizing agents (as examples: 70-90% formic acid, acetic acid [10%] pyridine, or relatively high concentrations of HCL (60 mM or more) are employed. Because such conditions would disrupt a tag/affinity matrix association, direct cleavage

of an affinity tag from the target protein while a protein remains associated with an affinity matrix is not attempted.

General conditions for cleavage at aspartate - proline sites are described in Current Protocols in Molecular Biology (supp. 28; chapter 16.4) John Wiley & Sons Inc. 1994, and in Landon, M. "Cleavage at Aspartyl - Prolyl Bonds" in Methods in Enzymology (1977) 47: 145-149. These references suggest that significant variability of cleavage conditions exist for different proteins and that cleavage might occur in some instances without first denaturing or solubilizing the protein. However, in practice, the latter circumstances are rare and proteins to be subjected to acid cleavage at Asp-Pro dipeptides are usually solubilized to a state where there is no visible turbidity. Such solubilized protein will normally not pellet when centrifuged at 100,000 x g for 1 hour. It is now shown that mild-acid conditions may be used for cleavage of aspartate-proline sites in Caulobacter S-layer fusion proteins without placing the protein in a solubilized state as described above.

## SUMMARY OF INVENTION

This invention is based on the unexpected discovery that recombinant fusion proteins produced by the Caulobacter S-layer protein secretion system can be cleaved under mild-acid conditions and solubilization of the fusion protein is not required. Cleavage may be accomplished while the fusion protein is in the form of an insoluble aggregate typical of the Caulobacter S-layer protein. Cleavage occurs at aspartate-protein dipeptides which may be in a heterologous protein portion of the fusion protein or in a portion that is native to the Caulobacter S-layer portion. The dipeptide may be placed at a desired location for cleavage by engineering DNA encoding the fusion protein to express the dipeptide at the desired location. A preferable location for cleavage may be at or near the junction between a heterologous (target) protein and the Caulobacter S-layer portion comprising the Caulobacter secretion signal, such that a cleavage product will be the target protein in its entirety and substantially free of extraneous amino acids.

The current invention makes it possible to cleave a heterologous (target) protein from the S-layer protein portion using only mild-acid conditions, even while the fusion protein is in an aggregated form. These cleavage conditions do not result in significant solubilization of the S-layer protein portion.

This invention provides a method of cleaving a fusion protein including a first component which comprises all or part of a Caulobacter S-layer protein including a Caulobacter C-terminal secretion signal, and a second component heterologous to Caulobacter. The fusion protein contains at least one aspartate-proline dipeptide. The method comprises combining the fusion protein with an acid solution of a strength insufficient to solubilize the fusion protein for a time sufficient for cleavage of the fusion protein at the aspartate-proline dipeptide. The acid solution may have a pH of from about 1.5 (eg. 1.5 ± 0.1) to about 2.5 (eg. 2.5 ± 0.1), and preferably from about 1.65 (eg. 1.65 ± 0.05) to about 2.35 (eg. 2.35 ± 0.05). Preferred pH conditions may be achieved using an acid equivalent in the range of about 5 to about 20 mM HCL. The method is typically carried out at a temperature in the range of approximately room temperature to about 50°C.

This invention also provides a method of preparing a DNA construct suitable for expression of a fusion protein suitable for use in the method of this invention. The method comprises joining an upstream DNA segment including DNA heterologous to Caulobacter which includes a protein of interest to a downstream DNA segment including DNA for a Caulobacter C-terminal secretion signal which does not encode an aspartate-proline dipeptide. The upstream segment contains DNA encoding an aspartate-proline dipeptide at or near the junction between said upstream and downstream segments.

This invention also provides a method of preparing a fusion protein, comprising the steps of expressing a DNA construct as described above in Caulobacter and recovering said fusion protein once secreted by the Caulobacter.

Once cleavage is accomplished according to this invention, the S-layer portion comprising the Caulobacter secretion signal may remain as an insoluble aggregate. If the target protein is soluble, the S-layer portion may be easily separated from the target

recombinant protein by simple centrifugation or filtration methods. Thus the system of this invention facilitates separation as would a Tag/affinity matrix system except that here, the system is also the means for producing an insoluble matrix. In addition, the insoluble matrix produced by this invention is resistant to the effects of the acid treatment, allowing direct cleavage of the target recombinant protein. In this way, a very inexpensive chemical cleavage method can be employed to economically retrieve recombinant proteins from a bacterial fusion protein. In contrast to the cost of most affinity matrices, there is little expense associated with the use of the S-layer secretion signal as it is simply a part of the fermentation/secretion process.

## DETAILED DESCRIPTION OF EMBODIMENTS OF THE INVENTION

### Production of Recombinant Fusion Proteins Using the Caulobacter S-layer Secretion System

Proteins may be produced using the Caulobacter S-layer Type 1 secretion pathway which requires only the C-terminal secretion signal of the Caulobacter. This signal is the C-terminal portion of the S-layer protein, which typically comprises about 200 amino acids. (See: Bingle, et al. (1997) [*supra*]; and, WO 97/34000). Additional Caulobacter S-layer DNA upstream from the secretion signal may also be present and may be desirable to encode portions of the S-layer protein which will contribute to aggregate formation of the secreted protein. Such additional Caulobacter DNA may constitute most or all of the remainder of the DNA encoding the S-layer protein.

Standard techniques (such as methods described in WO 97/34000) may be used to identify the amount of the C-terminal portion of a particular Caulobacter S-layer protein which functions as the secretion signal.

Creation of fusion proteins is commonly done by preparing DNA which codes for the target protein and fusing it in-frame with the C-terminal region of the S-layer gene. There are numerous possible methods, with the following being examples.

1. **Oligonucleotide Chemical Synthesis**. This involves the design of complementary single strands, complete with desirable restriction endonuclease cut sites

at the ends, chemical synthesis of the strands followed by annealing, cloning into a plasmid vector, juxtaposed to an appropriate portion of the C-terminal region of the S-layer gene.

2. **Production of the Target Gene DNA by Polymerase Chain Reaction (PCR) Amplification of a Target Sequence**. In this case, appropriate in-frame restriction sites are incorporated into the short oligonucleotides used for amplification of a target sequence, such that the final PCR product can be treated with the appropriate restriction enzymes (to create the restriction site "sticky ends"), followed by cloning into a plasmid vector, juxtaposed to an appropriate portion of the C-terminal region of the S-layer gene.

3. **Adapting Restriction Endonuclease Cleavage Sites that are Native to a Target Protein Gene Sequence for Fusion to the DNA Coding for the C-terminal S-layer Secretion Signal to Accomplish In-frame Expression of a Chimeric Protein.** This can be accomplished by direct ligation (although it is uncommon that an appropriate match will occur), or the use of adapter sequences or methods involving blunting of a restriction site and subsequent blunt-end ligation to change expression reading frame or join unlike restriction site sticky ends.

There will be numerous convenient sites for fusion with the C-terminal regions of the S-layer that lead to the successful expression, secretion and aggregation of a recombinant fusion protein. Some example positions are at or near the DNA sites corresponding to amino acids 622, 690, 784, 892 and 907 of the C. crescentus S-layer gene (see: Appendix 1 and, WO 97/34000). Other sites of fusion with the S-layer gene may also be employed. Most often a plasmid vector is designed such that the C-terminal gene segment is resident on a plasmid with appropriate restriction sites placed at the N-terminal junction of the S-layer fragment. Target recombinant protein gene segments are then cloned into those restriction sites. It is typical to prepare initial plasmid constructs that are replicated in E.coli. After a construct is produced, it is typically transferred to a broad host range plasmid which can then be introduced into the appropriate Caulobacter strain by electroporation. Suitable broad host range plasmids can be constructed from (but are not limited to) the IncQ, IncW and IncP1

plasmid incompatibility groups.

The introduction of the aspartate-proline (Asp-Pro) dipeptide at the appropriate site in the fusion protein can be done in several ways. Some examples are:

(a) incorporating a DNA sequence necessary to express the Asp-Pro dipeptide into the oligonucleotides used to prepare the target sequence, either by oligonucleotide synthesis or PCR methods;

(b) preparing a DNA segment with appropriate restriction sites at the termini so that an Asp-Pro dipeptide can be introduced (most often at the junction between S-layer and target gene) after a fusion recombinant S-layer gene has been made; and

(c) use of a native Asp-Pro dipeptide in either the target DNA or the S-layer segment (for example, an Asp-Pro dipeptide is located at amino acids 692 and 693 of the C. crescentus S-layer gene and is suitable for fusions made at the amino acid site).

The methods described above are not the only methods that may be used for creating and expressing fusion recombinant S-layer proteins, nor is it necessary to have the engineered genes resident on a plasmid. For example, the expressed gene may be introduced into the chromosome (using well-known gene insertion or replacement techniques) and still achieve secretion of the recombinant proteins (see WO 97/34000). In some cases it may be desirable to produce recombinant fusion proteins as insertions of heterologous DNA in the middle of the S-layer gene. In such a case, Asp-Pro dipeptide sequences could be engineered at the N and C-termini of the target peptide.

All possible codon combinations for Asp-Pro will work but the CCA codon for proline is not preferred due to the likelihood of a low amount of the corresponding tRNA being present in Caulobacter. The following is an approximate usage table for C. crescentus.

# TABLE 1

**Caulobacter crescentus Codon Usage Table**
[Amino Acid] [Triplet Code] [Frequency Per Thousand]

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Phe UUU | 2.5 | Ser UCU | 1.2 | Try UAU | 6.6 | Cys UGU | 0.6 |
| Phe UUC | 27.0 | Ser UCC | 8.5 | Try UAC | 9.6 | Cys UGC | 5.5 |
| Leu UUA | 0.0 | Ser UCA | 1.2 | STOP UAA | 0.8 | Cys UGA | 1.6 |
| Leu UUG | 4.4 | Ser UCG | 25.7 | STOP UAG | 0.6 | STOP UGG | 7.2 |
| Leu CUU | 4.4 | Pro CCU | 2.5 | His CAU | 3.2 | Arg CGU | 7.6 |
| Leu CUC | 15.7 | Pro CCC | 15.5 | His CAC | 12.2 | Arg CGC | 44.7 |
| Leu CUA | 1.1 | Pro CCA | 1.9 | Gln CAA | 3.7 | Arg CGA | 3.0 |
| Leu CUG | 72.3 | Pro CCG | 27.1 | Gln CAG | 30.2 | Arg CGG | 12.1 |
| IleAUU | 2.4 | Thr ACU | 1.2 | Asn AAU | 4.1 | Ser AGU | 0.8 |
| Ile AUC | 49.0 | Thr ACC | 37.3 | Asn AAC | 23.8 | Ser AGC | 14.9 |
| Ile AUA | 0.3 | Thr ACA | 0.8 | Lys AAA | 2.7 | Arg AGA | 0.4 |
| Met AUG | 25.7 | Thr ACG | 15.8 | Lys AAG | 37.9 | Arg AGG | 1.1 |
| Val GUU | 5.4 | Ala GCU | 9.5 | Asp GAU | 11.1 | Gly GGU | 9.5 |
| Val GUC | 42.7 | Ala GCC | 84.1 | Asp GAC | 48.5 | Gly GGC | 64.8 |
| Val GUA | 1.0 | Ala GCA | 2.2 | Glu GAA | 20.5 | Gly GGA | 2.3 |
| Val GUG | 30.7 | Ala GCG | 36.7 | Glu GAG | 45.4 | Gly GGG | 7.7 |

Large quantities (eg. 12% of total cell protein/3% of input organic carbon) of a wide range of proteins can be produced, with yields in the order of 250 mg/liter of batch culture. Fusion proteins with 35 kDa of target peptide are secreted with little difficulty, although proteins with multiple cysteines may be more difficult to express. Post-expression glycosylation of proteins does not occur, an advantage for most peptide expression applications.

**Host Expression Strains**

For secretion of recombinant fusion S-layer proteins, the Caulobacter strain will preferably be one which has lost the ability to produce a native S-layer protein, while retaining a fully functional S-layer protein secretion apparatus. Such strains may be obtained by screening for mutants that have spontaneously become S-layer protein negative; or, by directed genetic manipulation, such as (but not limited to) the insertion of a drug resistance cassette in the middle of the S-layer gene or the substitution of a version of the S-layer gene which has had a sizeable internal region deleted from the gene (see: Bingle et al. 1997[1] [*supra*]; Bingle et al. 1997[2] "Cell Surface Display of a Pseudonomonas aerugenosa PAK Pilin Peptide with the Paracrystalline Layer of Caulobacter crescentus" Molec. Microbiol. 26:277-288; and, Edwards and Smit (1991) " A Transducing Bacteriophage for Caulobacter us Uses the Paracrystalline Surface Layer Protein as a Receptor" J. Bacteriol. 173: 5568-5572). In the case of a genetic manipulation, a common method for producing such strains is to modify a copy of the S-layer gene while on a plasmid and then to use well known gene replacement methods to substitute the modified gene for the native gene in the Caulobacter chromosome (see: Edwards and Smit (1991) [*supra*]).

If an entire S-layer gene is to be used for production of a recombinant protein (via insertion of a target sequence), strains defective in the production of the lipopolysacharide (LPS) used for S-layer attachment to the bacterial surface can be used. These can be prepared by forcing Caulobacter to grow without exogenous

calcium. Under these conditions mutants arise that are uniformly defective in producing a proficient version of the S-layer LPS (see: Walker, S.G. et al. (1994) "Characteristics of Mutants of Caulobacter crescentus Defective in Surface Attachment of the Paracrystaline Layer" J. Bacteriol. 176: 6312-6323).

All Caulobacter S-layer producing strains are suitable for this technology. One may isolate the S-layer gene from a particular strain (using homology between Caulobacter S-layers to design probes to detect and clone the S-layer genes) and adapt the C-terminal region for recombinant protein expression, in a manner similar to that done for C. crescentus strains (see: MacRae and Smit (1991) [*supra*], and Walker. S.G. et al. (1992) [*supra*]). Alternatively, one may construct recombinant fusion S-layer genes using the C. crescentus S-layer gene and express the recombinant genes in alternate Caulobacter hosts.

Freshwater Caulobacter producing S-layers may be readily detected by negative stain transmission electron microscopy techniques. Caulobacter may be isolated using the methods outlined by MacRae and Smit (1991) [*supra*], which take advantage of the fact that Caulobacter can tolerate periods of starvation while other soil and water bacteria may not and that they all produce a distinctive stalk structure, visible by light microscopy (using either phase contrast or standard dye staining methods). Once Caulobacter strains are isolated in a typical procedure, colonies may be suspended in 2% ammonium molybdate negative stain and applied to plastic-filmed, carbon-stabilized 300 or 400 mesh copper or nickel grids and examined in a transmission electron microscope at 60 kilovolt accelerating voltage (see: Smit, J. (1986) "Protein Surface Layers of Bacteria", in Outer Membranes as Model Systems, (M. Inouge, ed. J.Wiley & Sons, at p. 343-376). S-layers are seen as two-dimensional geometric patterns most readily on those cells in a colony that have lysed and released their internal contents.

**Recombinant Protein Purification**

Secreted proteins are separated and shed into the culture media as a macroscopic precipitate (the "aggregate" referred to herein). The shedding phenomenon is a consequence of the absence of the N-terminal region of the S-layer protein in the

expressed recombinant protein, or the loss of the lipopolysaccharide species used for S-layer attachment by the Caulobacter (see: Walker, S.G. et al. (1994) [*supra*]). Typically, the aggregate forms as loose, gel-like lumps of pure protein that can readily be retrieved and separated from the bacteria by simple filtration.

The aggregate may be readily separated from a soluble cleaved target protein by any suitable techniques such as filtration of centrifugation. If the target protein is insoluble once cleaved, it may then be convenient to then solubilize one or both of the proteins (for example in 8M urea or 6M quanidine HCL) and separate by chromatography. In this way, only 2 species of protein need to be separated.

**Cleavage of Fusion Proteins**

General procedures for performing mild-acid cleavage are known from in the prior art as described above. In the method of this invention, conditions are adjusted to avoid destruction of the target protein or solubilization of the aggregate containing the S-layer secretion signal. Excess acid or too high a temperature may increase the occurrence over time of random cleavages along the length of the fusion protein, which is to be avoided since such random cleavages may lead to undersized fragmentation of the fusion protein or solubilization of the aggregated S-layer portion.

Good yields of target protein with minimum random breaks in the fusion protein may generally be achieved by using from 5-20 mM HCL (or its equivalent while employing another acid). The respective pH of these conditions (unbuffered acid solution) is from about 2.3 to about 1.7. Time and temperature is preferably adjusted by routine monitoring to achieve the desired cleavage while minimizing random breaks. For example, temperature may range from room temperature to about $50^0$ C. Time of treatment may range from about 12 to about 72 hours. Time or temperature outside of these ranges is permissible depending upon the strength of the acid and the accepted yield. Generally, lower yields are obtained with less acid strength, less time or lower temperatures.

In the following examples, efficiency of cleavage in the order of 40-80% is

achieved using conditions the same as or similar to the following alternatives:

- 5 mM HCL at $50^0$ C. for 48-72 hours

- 20 mM HCL at $30^0$ C. for 48-72 hours.

Conditions in excess of the aforementioned values may be employed in some cases with the possibility of random breaks increasing, particularly with increased acid strength or temperature. In the following examples, significant random cleavage occurred with 50 mM HCL at $50^0$ C. after 48 hours.

Any acid may be employed in this invention which is normally used in solutions to which proteins are exposed. Acids which have a deleterious effect on proteins under dilute conditions should be avoided. For example, HCL or an equivalent amount of $H_2SO_4$ may be used in this invention but oxidizing acids such as nitric acid may not be suitable.

**Example 1. Cleavage of artificial silk protein sequences from a secretion signal containing a native aspartate-proline cleavage site.**

An artificial protein sequence resembling spider silk was constructed by synthesis of partially overlapping and complementing oligomers of DNA, which were then completed to a full duplex DNA with Taql polymerase extension, to create a sequence that coded for 97 amino acids. The resulting DNA sequence and corresponding amino acid sequence are shown in Appendix 2.

The DNA sequence shown in Appendix 2 was cloned into a gene carrier sequence residing in a pUC8 plasmid cloning vector. The gene segment carrier had BamH1 restriction sites at each end and an internal BglII site. This combination of restrictions sites allowed the production of multimers of the above sequence, relying on the fact that BamH1 sticky ends will ligate into BglII sticky end, with the loss of both restriction sites. Thus one copy of the silk-like sequence within the gene segment carrier can be put inside a second copy of the same to produce a dimer. Using this principle, an 8X repeat was produced, fused to DNA encoding the S-layer secretion signal corresponding to the C-terminal portion of the C. crescentus S-layer protein from about amino acid 690 onwards (see: Appendix 1). This fusion protein gene was

introduced into strain CB2A on a broad host range plasmid vector. The 8x multimer appeared to be unstable, resulting in recombination events that reduced the 8X multimer to a 3x size. The 3 fold repeat of the above 97 amino acid sequence, fused to the S-layer secretion signal was secreted. Protein was collected and subjected to treatment with 5mM HCL for 2 days at 50° C. The result was the liberation of about 80% of soluble silk-like polymer which was readily separated by filtration from the S-layer protein which remained completely aggregated under these conditions. Cleavage occurred at native aspartate-proline dimer in the Caulobacter S-layer signal region (see: Appendix 1, amino acids numbered 692-693).

**Example 2. Cleavage of the salmonid virus Infectious Pancreatic Necrosis Virus (IPNV) surface glycoprotein candidate vaccine sequence from an S-layer secretion signal containing a native aspartate-proline site.**

The surface glycoprotein of the IPNV strain is a vaccine candidate. For this example and Example 4, the sequence of the first 257 amino acids of the mature protein and the corresponding DNA sequence as shown in Appendix 3 were used.

DNA encoding a segment of the major surface glycoprotein gene of IPNV specifying amino acids 145-257 of the protein was fused to DNA sequence specifying two putative T-cell activating epitopes: MVF (SEQ ID No:1; LSEIKGVIVHRLEGV, derived from Measles Virus protein F) and P2 (SEQ ID No:2; QYIKANSKFIGITEL, derived from tetanus toxoid protein). The T-cell epitopes were positioned on the C-terminal end of the IPNV sequence. This chimeric protein was in turn fused in frame with the C-crescentus S-layer gene at about amino acid 690 position of the gene and introduced into Caulobacter on a broad host range plasmid vector. The resulting secreted protein was collected and treated with 5 mM HCL for 2 days at 50° C. Cleavage occurred at the native aspartate-proline dimer described in Example 1. The result was the liberation of about 75% of soluble vaccine candidate chimeric protein from the S-layer secretion signal which remained aggregated.

**Example 3. Cleavage of segments of an E. coli type I pilus tip subunit from an S-layer secretion signal containing a native aspartate-proline cleavage site.**

The FimH gene product is the tip pilus subunit of the E. coli strains involved with urinary tract infections. Two segments, T3 (specifying the first 145 amino acids of the mature peptide) and T7 (specifying the entire 258 amino acids of the mature peptide) were fused to the S-layer secretion signal at about amino acid 690 of the S-layer sequence. The T3 and T7 sequences are shown in Appendix 4.

The fusion protein genes were introduced into strain CB2A on a broad host range plasmid vector. In both cases the resulting secreted protein was collected and treated with 5 mM HCL for 2 days at 50° C. In both cases, the result was the liberation of about 50% of soluble vaccine candidate chimeric protein from the S-layer secretion signal which remained aggregated. Cleavage occurred at the native aspartate-proline dimer described in Example 1.

**Example 4. Cleavage of the salmonid virus IPNV surface glycoprotein candidate vaccine sequence from an S-layer secretion signal containing an introduced aspartate-proline cleavage site.**

A segment of the major surface glycoprotein gene of IPNV specifying amino acids 1-257 of the protein shown in Appendix 4 was fused to a DNA sequence specifying a peptide containing an aspartate-proline dipeptide (SEQ ID No: 3; SPLGPAGDPEAS) such that the aspartate-proline dipeptide was positioned very near the C-terminus of the chimeric protein. This chimeric protein was in turn fused in frame with the C. crescentus S-layer gene at about amino acid 784 position of the gene and introduced in strain CB2A on a broad host range plasmid vector. The resulting secreted protein was collected and treated with 5 mM HCL for 2 days at 50° C. Cleavage occurred at the introduced aspartate-proline dipeptide. The result was the liberation of about 40% of insoluble vaccine candidate chimeric protein from the S-layer secretion signal which remained aggregated.

Longer DNA and amino acid sequences referred to above are set out in the

following Appendices which are part of this description. Appendix 1 sets out the complete nucleotide sequence of the C. crescentus S-layer gene (SEQ ID No: 4) with the upstream sequence including the -35 and -10 sites of the promoter region and the Shine Dalgarno sequence. The start codon is at nucleotide 101 and the coding sequence run to and includes nucleotide 3179. The amino acid sequence of the C. crescentus S-layer protein (SEQ ID No: 5) included in Appendix 1 is predicted from the DNA sequence. Appendix 2 sets out the artificial spider silk DNA sequence (SEQ ID No:6) used in Example 1 and the corresponding amino acid sequence (SEQ ID No. 7). Appendix 3 sets out the DNA sequence (SEQ ID No: 8) and corresponding amino acid sequence (SEQ ID No: 9) of the first 257 amino acids of IPNV as described in Examples 2 and 4. Appendix 4 sets out the T3 protein sequence (SEQ ID No: 10) and the T7 protein sequence (SEQ ID No: 11) as described in Example 3.

All publications, patents and patent applications referred to herein are hereby incorporated by reference. While this invention has been described according to particular embodiments and by reference to certain examples, it will be apparent to those of skill in the art that variations and modifications of the invention as described herein fall within the spirit and scope of the attached claims.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant : John Smit
Serial No. : 09/743,731
Filed : January 12, 2001
Title : CLEAVAGE OF CAULOBACTER PRODUCED RECOMBINANT FUSION PROTEINS

Art Unit : Unknown
Examiner : Unknown

**Box PCT**
Commissioner for Patents
Washington, D.C. 20231

VERIFIED STATEMENT UNDER 37 CFR §1.821(f)

I, Katica Magovcevic, declare that I personally prepared the paper and the computer-readable copy of the Sequence Listing filed herewith for the above-identified application and that the content of both is the same.

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of The United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Date: 4/25/01

Katica Magovcevic

Fish & Richardson P.C.
225 Franklin Street
Boston, MA 02110-2804
(617) 542-5070 telephone
(617) 542-8906 facsimile

20208429.doc

CERTIFICATE OF MAILING BY EXPRESS MAIL

Express Mail Label No. EL259011499US

I hereby certify under 37 CFR §1.10 that this correspondence is being deposited with the United States Postal Service as Express Mail Post Office to Addressee with sufficient postage on the date indicated below and is addressed to the Commissioner for Patents, Washington, D.C. 20231.

Date of Deposit: April 25, 2001

Signature: Samantha Bell

Typed or Printed Name of Person Signing Certificate: Samantha Bell

09/743731





SEQUENCE LISTING

<110> Smit, John

<120> CLEAVAGE OF CAULOBACTER PRODUCED
RECOMBINANT FUSION PROTEINS

<130> 08106-004001

<140> 09/743,731
<141> 2001-01-12

<150> PCT/CA99/00637
<151> 1999-07-14

<150> CA 2,237,704
<151> 1998-07-14

<160> 11

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Fusion protein

<400> 1
Leu Ser Glu Ile Lys Gly Val Ile Val His Arg Leu Glu Gly Val
1 5 10 15

<210> 2
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Fusion protein

<400> 2
Gln Tyr Ile Lys Ala Asn Ser Lys Phe Ile Gly Ile Thr Glu Leu
1 5 10 15

<210> 3
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetically generated peptide

<400> 3
Ser Pro Leu Gly Pro Ala Gly Asp Pro Glu Ala Ser

```
 1               5                   10

<210> 4
<211> 3300
<212> DNA
<213> Caulobacter crescentus

<220>
<221> CDS
<222> (101)...(3179)

<400> 4
gctattgtcg acgtatgacg tttgctctat agccatcgct gctcccatgc gcgccactcg      60
gtcgcagggg gtgtgggatt ttttttggga gacaatcctc atg gcc tat acg acg       115
                                            Met Ala Tyr Thr Thr
                                             1               5

gcc cag ttg gtg act gcg tac acc aac gcc aac ctc ggc aag gcg cct       163
Ala Gln Leu Val Thr Ala Tyr Thr Asn Ala Asn Leu Gly Lys Ala Pro
                10                  15                  20

gac gcc gcc acc acg ctg acg ctc gac gcg tac gcg act caa acc cag       211
Asp Ala Ala Thr Thr Leu Thr Leu Asp Ala Tyr Ala Thr Gln Thr Gln
            25                  30                  35

acg ggc ggc ctc tcg gac gcc gct gcg ctg acc aac acc ctg aag ctg       259
Thr Gly Gly Leu Ser Asp Ala Ala Ala Leu Thr Asn Thr Leu Lys Leu
        40                  45                  50

gtc aac agc acg acg gct gtt gcc atc cag acc tac cag ttc ttc acc       307
Val Asn Ser Thr Thr Ala Val Ala Ile Gln Thr Tyr Gln Phe Phe Thr
    55                  60                  65

ggc gtt gcc ccg tcg gcc gct ggt ctg gac ttc ctg gtc gac tcg acc       355
Gly Val Ala Pro Ser Ala Ala Gly Leu Asp Phe Leu Val Asp Ser Thr
70                  75                  80                  85

acc aac acc aac gac ctg aac gac gcg tac tac tcg aag ttc gct cag       403
Thr Asn Thr Asn Asp Leu Asn Asp Ala Tyr Tyr Ser Lys Phe Ala Gln
                90                  95                  100

gaa aac cgc ttc atc aac ttc tcg atc aac ctg gcc acg ggc gcc ggc       451
Glu Asn Arg Phe Ile Asn Phe Ser Ile Asn Leu Ala Thr Gly Ala Gly
            105                 110                 115

gcc ggc gcg acg gct ttc gcc gcc gcc tac acg ggc gtt tcg tac gcc       499
Ala Gly Ala Thr Ala Phe Ala Ala Ala Tyr Thr Gly Val Ser Tyr Ala
        120                 125                 130

cag acg gtc gcc acc gcc tat gac aag atc atc ggc aac gcc gtc gcg       547
Gln Thr Val Ala Thr Ala Tyr Asp Lys Ile Ile Gly Asn Ala Val Ala
    135                 140                 145

acc gcc gct ggc gtc gac gtc gcg gcc gcc gtg gct ttc ctg agc cgc       595
Thr Ala Ala Gly Val Asp Val Ala Ala Ala Val Ala Phe Leu Ser Arg
150                 155                 160                 165
```

```
cag gcc aac atc gac tac ctg acc gcc ttc gtg cgc gcc aac acg ccg      643
Gln Ala Asn Ile Asp Tyr Leu Thr Ala Phe Val Arg Ala Asn Thr Pro
                170                 175                 180

ttc acg gcc gct gcc gac atc gat ctg gcc gtc aag gcc gcc ctg atc      691
Phe Thr Ala Ala Ala Asp Ile Asp Leu Ala Val Lys Ala Ala Leu Ile
            185                 190                 195

ggc acc atc ctg aac gcc gcc acg gtg tcg ggc atc ggt ggt tac gcg      739
Gly Thr Ile Leu Asn Ala Ala Thr Val Ser Gly Ile Gly Gly Tyr Ala
        200                 205                 210

acc gcc acg gcc gcg atg atc aac gac ctg tcg gac ggc gcc ctg tcg      787
Thr Ala Thr Ala Ala Met Ile Asn Asp Leu Ser Asp Gly Ala Leu Ser
    215                 220                 225

acc gac aac gcg gct ggc gtg aac ctg ttc acc gcc tat ccg tcg tcg      835
Thr Asp Asn Ala Ala Gly Val Asn Leu Phe Thr Ala Tyr Pro Ser Ser
230                 235                 240                 245

ggc gtg tcg ggt tcg acc ctc tcg ctg acc acc ggc acc gac acc ctg      883
Gly Val Ser Gly Ser Thr Leu Ser Leu Thr Thr Gly Thr Asp Thr Leu
                250                 255                 260

acg ggc acc gcc aac aac gac acg ttc gtt gcg ggt gaa gtc gcc ggc      931
Thr Gly Thr Ala Asn Asn Asp Thr Phe Val Ala Gly Glu Val Ala Gly
            265                 270                 275

gct gcg acc ctg acc gtt ggc gac acc ctg agc ggc ggt gct ggc acc      979
Ala Ala Thr Leu Thr Val Gly Asp Thr Leu Ser Gly Gly Ala Gly Thr
        280                 285                 290

gac gtc ctg aac tgg gtg caa gct gct gcg gtt acg gct ctg ccg acc     1027
Asp Val Leu Asn Trp Val Gln Ala Ala Ala Val Thr Ala Leu Pro Thr
    295                 300                 305

ggc gtg acg atc tcg ggc atc gaa acg atg aac gtg acg tcg ggc gct     1075
Gly Val Thr Ile Ser Gly Ile Glu Thr Met Asn Val Thr Ser Gly Ala
310                 315                 320                 325

gcg atc acc ctg aac acg tct tcg ggc gtg acg ggt ctg acc gcc ctg     1123
Ala Ile Thr Leu Asn Thr Ser Ser Gly Val Thr Gly Leu Thr Ala Leu
                330                 335                 340

aac acc aac acc agc ggc gcg gct caa acc gtc acc gcc ggc gct ggc     1171
Asn Thr Asn Thr Ser Gly Ala Ala Gln Thr Val Thr Ala Gly Ala Gly
            345                 350                 355

cag aac ctg acc gcc acg acc gcc gct caa gcc gcg aac aac gtc gcc     1219
Gln Asn Leu Thr Ala Thr Thr Ala Ala Gln Ala Ala Asn Asn Val Ala
        360                 365                 370

gtc gac ggg cgc gcc aac gtc acc gtc gcc tcg acg ggc gtg acc tcg     1267
Val Asp Gly Arg Ala Asn Val Thr Val Ala Ser Thr Gly Val Thr Ser
    375                 380                 385

ggc acg acc acg gtc ggc gcc aac tcg gcc gct tcg ggc acc gtg tcg     1315
```

Gly Thr Thr Thr Val Gly Ala Asn Ser Ala Ala Ser Gly Thr Val Ser
390 395 400 405

gtg agc gtc gcg aac tcg agc acg acc acc acg ggc gct atc gcc gtg 1363
Val Ser Val Ala Asn Ser Ser Thr Thr Thr Thr Gly Ala Ile Ala Val
410 415 420

acc ggt ggt acg gcc gtg acc gtg gct caa acg gcc ggc aac gcc gtg 1411
Thr Gly Gly Thr Ala Val Thr Val Ala Gln Thr Ala Gly Asn Ala Val
425 430 435

aac acc acg ttg acg caa gcc gac gtg acc gtg acc ggt aac tcc agc 1459
Asn Thr Thr Leu Thr Gln Ala Asp Val Thr Val Thr Gly Asn Ser Ser
440 445 450

acc acg gcc gtg acg gtc acc caa acc gcc gcc gcc acc gcc ggc gct 1507
Thr Thr Ala Val Thr Val Thr Gln Thr Ala Ala Ala Thr Ala Gly Ala
455 460 465

acg gtc gcc ggt cgc gtc aac ggc gct gtg acg atc acc gac tct gcc 1555
Thr Val Ala Gly Arg Val Asn Gly Ala Val Thr Ile Thr Asp Ser Ala
470 475 480 485

gcc gcc tcg gcc acg acc gcc ggc aag atc gcc acg gtc acc ctg ggc 1603
Ala Ala Ser Ala Thr Thr Ala Gly Lys Ile Ala Thr Val Thr Leu Gly
490 495 500

agc ttc ggc gcc gcc acg atc gac tcg agc gct ctg acg acc gtc aac 1651
Ser Phe Gly Ala Ala Thr Ile Asp Ser Ser Ala Leu Thr Thr Val Asn
505 510 515

ctg tcg ggc acg ggc acc tcg ctc ggc atc ggc cgc ggc gct ctg acc 1699
Leu Ser Gly Thr Gly Thr Ser Leu Gly Ile Gly Arg Gly Ala Leu Thr
520 525 530

gcc acg ccg acc gcc aac acc ctg acc ctg aac gtc aat ggt ctg acg 1747
Ala Thr Pro Thr Ala Asn Thr Leu Thr Leu Asn Val Asn Gly Leu Thr
535 540 545

acg acc ggc gcg atc acg gac tcg gaa gcg gct gct gac gat ggt ttc 1795
Thr Thr Gly Ala Ile Thr Asp Ser Glu Ala Ala Ala Asp Asp Gly Phe
550 555 560 565

acc acc atc aac atc gct ggt tcg acc gcc tct tcg acg atc gcc agc 1843
Thr Thr Ile Asn Ile Ala Gly Ser Thr Ala Ser Ser Thr Ile Ala Ser
570 575 580

ctg gtg gcc gcc gac gcg acg acc ctg aac atc tcg ggc gac gct cgc 1891
Leu Val Ala Ala Asp Ala Thr Thr Leu Asn Ile Ser Gly Asp Ala Arg
585 590 595

gtc acg atc acc tcg cac acc gct gcc gcc ctg acg ggc atc acg gtg 1939
Val Thr Ile Thr Ser His Thr Ala Ala Ala Leu Thr Gly Ile Thr Val
600 605 610

acc aac agc gtt ggt gcg acc ctc ggc gcc gaa ctg gcg acc ggt ctg 1987
Thr Asn Ser Val Gly Ala Thr Leu Gly Ala Glu Leu Ala Thr Gly Leu

615 620 625

gtc ttc acg ggc ggc gct ggc cgt gac tcg atc ctg ctg ggc gcc acg 2035
Val Phe Thr Gly Gly Ala Gly Arg Asp Ser Ile Leu Leu Gly Ala Thr
630 635 640 645

acc aag gcg atc gtc atg ggc gcc ggc gac gac acc gtc acc gtc agc 2083
Thr Lys Ala Ile Val Met Gly Ala Gly Asp Asp Thr Val Thr Val Ser
650 655 660

tcg gcg acc ctg ggc gct ggt ggt tcg gtc aac ggc ggc gac ggc acc 2131
Ser Ala Thr Leu Gly Ala Gly Gly Ser Val Asn Gly Gly Asp Gly Thr
665 670 675

gac gtt ctg gtg gcc aac gtc aac ggt tcg tcg ttc agc gct gac ccg 2179
Asp Val Leu Val Ala Asn Val Asn Gly Ser Ser Phe Ser Ala Asp Pro
680 685 690

gcc ttc ggc ggc ttc gaa acc ctc cgc gtc gct ggc gcg gcg gct caa 2227
Ala Phe Gly Gly Phe Glu Thr Leu Arg Val Ala Gly Ala Ala Ala Gln
695 700 705

ggc tcg cac aac gcc aac ggc ttc acg gct ctg caa ctg ggc gcg acg 2275
Gly Ser His Asn Ala Asn Gly Phe Thr Ala Leu Gln Leu Gly Ala Thr
710 715 720 725

gcg ggt gcg acg acc ttc acc aac gtt gcg gtg aat gtc ggc ctg acc 2323
Ala Gly Ala Thr Thr Phe Thr Asn Val Ala Val Asn Val Gly Leu Thr
730 735 740

gtt ctg gcg gct ccg acc ggt acg acg acc gtg acc ctg gcc aac gcc 2371
Val Leu Ala Ala Pro Thr Gly Thr Thr Thr Val Thr Leu Ala Asn Ala
745 750 755

acg ggc acc tcg gac gtg ttc aac ctg acc ctg tcg tcc tcg gcc gct 2419
Thr Gly Thr Ser Asp Val Phe Asn Leu Thr Leu Ser Ser Ser Ala Ala
760 765 770

ctg gcc gct ggt acg gtt gcg ctg gct ggc gtc gag acg gtg aac atc 2467
Leu Ala Ala Gly Thr Val Ala Leu Ala Gly Val Glu Thr Val Asn Ile
775 780 785

gcc gcc acc gac acc aac acg acc gct cac gtc gac acg ctg acg ctg 2515
Ala Ala Thr Asp Thr Asn Thr Thr Ala His Val Asp Thr Leu Thr Leu
790 795 800 805

caa gcc acc tcg gcc aag tcg atc gtg gtg acg ggc aac gcc ggt ctg 2563
Gln Ala Thr Ser Ala Lys Ser Ile Val Val Thr Gly Asn Ala Gly Leu
810 815 820

aac ctg acc aac acc ggc aac acg gct gtc acc agc ttc gac gcc agc 2611
Asn Leu Thr Asn Thr Gly Asn Thr Ala Val Thr Ser Phe Asp Ala Ser
825 830 835

gcc gtc acc ggc acg gct ccg gct gtg acc ttc gtg tcg gcc aac acc 2659
Ala Val Thr Gly Thr Ala Pro Ala Val Thr Phe Val Ser Ala Asn Thr
840 845 850

acg gtg ggt gaa gtc gtc acg atc cgc ggc ggc gct ggc gcc gac tcg 2707
Thr Val Gly Glu Val Val Thr Ile Arg Gly Gly Ala Gly Ala Asp Ser
855 860 865

ctg acc ggt tcg gcc acc gcc aat gac acc atc atc ggt ggc gct ggc 2755
Leu Thr Gly Ser Ala Thr Ala Asn Asp Thr Ile Ile Gly Gly Ala Gly
870 875 880 885

gct gac acc ctg gtc tac acc ggc ggt acg gac acc ttc acg ggt ggc 2803
Ala Asp Thr Leu Val Tyr Thr Gly Gly Thr Asp Thr Phe Thr Gly Gly
890 895 900

acg ggc gcg gat atc ttc gat atc aac gct atc ggc acc tcg acc gct 2851
Thr Gly Ala Asp Ile Phe Asp Ile Asn Ala Ile Gly Thr Ser Thr Ala
905 910 915

ttc gtg acg atc acc gac gcc gct gtc ggc gac aag ctc gac ctc gtc 2899
Phe Val Thr Ile Thr Asp Ala Ala Val Gly Asp Lys Leu Asp Leu Val
920 925 930

ggc atc tcg acg aac ggc gct atc gct gac ggc gcc ttc ggc gct gcg 2947
Gly Ile Ser Thr Asn Gly Ala Ile Ala Asp Gly Ala Phe Gly Ala Ala
935 940 945

gtc acc ctg ggc gct gct gcg acc ctg gct cag tac ctg gac gct gct 2995
Val Thr Leu Gly Ala Ala Ala Thr Leu Ala Gln Tyr Leu Asp Ala Ala
950 955 960 965

gct gcc ggc gac ggc agc ggc acc tcg gtt gcc aag tgg ttc cag ttc 3043
Ala Ala Gly Asp Gly Ser Gly Thr Ser Val Ala Lys Trp Phe Gln Phe
970 975 980

ggc ggc gac acc tat gtc gtc gtt gac agc tcg gct ggc gcg acc ttc 3091
Gly Gly Asp Thr Tyr Val Val Val Asp Ser Ser Ala Gly Ala Thr Phe
985 990 995

gtc agc ggc gct gac gcg gtg atc aag ctg acc ggt ctg gtc acg ctg 3139
Val Ser Gly Ala Asp Ala Val Ile Lys Leu Thr Gly Leu Val Thr Leu
1000 1005 1010

acc acc tcg gcc ttc gcc acc gaa gtc ctg acg ctc gcc t aagcgaacgt 3189
Thr Thr Ser Ala Phe Ala Thr Glu Val Leu Thr Leu Ala
1015 1020 1025

ctgatcctcg cctaggcgag gatcgctaga ctaagagacc ccgtcttccg aaagggaggc 3249
ggggtctttc ttatgggcgc tacgcgctgg ccggccttgc ctagttccgg t 3300

<210> 5
<211> 1026
<212> PRT
<213> Caulobacter crescentus

<400> 5
Met Ala Tyr Thr Thr Ala Gln Leu Val Thr Ala Tyr Thr Asn Ala Asn
1 5 10 15
Leu Gly Lys Ala Pro Asp Ala Ala Thr Thr Leu Thr Leu Asp Ala Tyr

20 25 30
Ala Thr Gln Thr Gln Thr Gly Gly Leu Ser Asp Ala Ala Ala Leu Thr
35 40 45
Asn Thr Leu Lys Leu Val Asn Ser Thr Thr Ala Val Ala Ile Gln Thr
50 55 60
Tyr Gln Phe Phe Thr Gly Val Ala Pro Ser Ala Ala Gly Leu Asp Phe
65 70 75 80
Leu Val Asp Ser Thr Thr Asn Thr Asn Asp Leu Asn Asp Ala Tyr Tyr
85 90 95
Ser Lys Phe Ala Gln Glu Asn Arg Phe Ile Asn Phe Ser Ile Asn Leu
100 105 110
Ala Thr Gly Ala Gly Ala Gly Ala Thr Ala Phe Ala Ala Ala Tyr Thr
115 120 125
Gly Val Ser Tyr Ala Gln Thr Val Ala Thr Ala Tyr Asp Lys Ile Ile
130 135 140
Gly Asn Ala Val Ala Thr Ala Ala Gly Val Asp Val Ala Ala Ala Val
145 150 155 160
Ala Phe Leu Ser Arg Gln Ala Asn Ile Asp Tyr Leu Thr Ala Phe Val
165 170 175
Arg Ala Asn Thr Pro Phe Thr Ala Ala Ala Asp Ile Asp Leu Ala Val
180 185 190
Lys Ala Ala Leu Ile Gly Thr Ile Leu Asn Ala Ala Thr Val Ser Gly
195 200 205
Ile Gly Gly Tyr Ala Thr Ala Thr Ala Ala Met Ile Asn Asp Leu Ser
210 215 220
Asp Gly Ala Leu Ser Thr Asp Asn Ala Ala Gly Val Asn Leu Phe Thr
225 230 235 240
Ala Tyr Pro Ser Ser Gly Val Ser Gly Ser Thr Leu Ser Leu Thr Thr
245 250 255
Gly Thr Asp Thr Leu Thr Gly Thr Ala Asn Asn Asp Thr Phe Val Ala
260 265 270
Gly Glu Val Ala Gly Ala Ala Thr Leu Thr Val Gly Asp Thr Leu Ser
275 280 285
Gly Gly Ala Gly Thr Asp Val Leu Asn Trp Val Gln Ala Ala Ala Val
290 295 300
Thr Ala Leu Pro Thr Gly Val Thr Ile Ser Gly Ile Glu Thr Met Asn
305 310 315 320
Val Thr Ser Gly Ala Ala Ile Thr Leu Asn Thr Ser Ser Gly Val Thr
325 330 335
Gly Leu Thr Ala Leu Asn Thr Asn Thr Ser Gly Ala Ala Gln Thr Val
340 345 350
Thr Ala Gly Ala Gly Gln Asn Leu Thr Ala Thr Thr Ala Ala Gln Ala
355 360 365
Ala Asn Asn Val Ala Val Asp Gly Arg Ala Asn Val Thr Val Ala Ser
370 375 380
Thr Gly Val Thr Ser Gly Thr Thr Thr Val Gly Ala Asn Ser Ala Ala
385 390 395 400
Ser Gly Thr Val Ser Val Ser Val Ala Asn Ser Ser Thr Thr Thr Thr
405 410 415
Gly Ala Ile Ala Val Thr Gly Gly Thr Ala Val Thr Val Ala Gln Thr
420 425 430
Ala Gly Asn Ala Val Asn Thr Thr Leu Thr Gln Ala Asp Val Thr Val
435 440 445
Thr Gly Asn Ser Ser Thr Thr Ala Val Thr Val Thr Gln Thr Ala Ala
450 455 460
Ala Thr Ala Gly Ala Thr Val Ala Gly Arg Val Asn Gly Ala Val Thr
465 470 475 480

Ile Thr Asp Ser Ala Ala Ala Ser Ala Thr Thr Ala Gly Lys Ile Ala
485 490 495
Thr Val Thr Leu Gly Ser Phe Gly Ala Ala Thr Ile Asp Ser Ser Ala
500 505 510
Leu Thr Thr Val Asn Leu Ser Gly Thr Gly Thr Ser Leu Gly Ile Gly
515 520 525
Arg Gly Ala Leu Thr Ala Thr Pro Thr Ala Asn Thr Leu Thr Leu Asn
530 535 540
Val Asn Gly Leu Thr Thr Thr Gly Ala Ile Thr Asp Ser Glu Ala Ala
545 550 555 560
Ala Asp Asp Gly Phe Thr Thr Ile Asn Ile Ala Gly Ser Thr Ala Ser
565 570 575
Ser Thr Ile Ala Ser Leu Val Ala Ala Asp Ala Thr Thr Leu Asn Ile
580 585 590
Ser Gly Asp Ala Arg Val Thr Ile Thr Ser His Thr Ala Ala Ala Leu
595 600 605
Thr Gly Ile Thr Val Thr Asn Ser Val Gly Ala Thr Leu Gly Ala Glu
610 615 620
Leu Ala Thr Gly Leu Val Phe Thr Gly Gly Ala Gly Arg Asp Ser Ile
625 630 635 640
Leu Leu Gly Ala Thr Thr Lys Ala Ile Val Met Gly Ala Gly Asp Asp
645 650 655
Thr Val Thr Val Ser Ser Ala Thr Leu Gly Ala Gly Gly Ser Val Asn
660 665 670
Gly Gly Asp Gly Thr Asp Val Leu Val Ala Asn Val Asn Gly Ser Ser
675 680 685
Phe Ser Ala Asp Pro Ala Phe Gly Gly Phe Glu Thr Leu Arg Val Ala
690 695 700
Gly Ala Ala Ala Gln Gly Ser His Asn Ala Asn Gly Phe Thr Ala Leu
705 710 715 720
Gln Leu Gly Ala Thr Ala Gly Ala Thr Thr Phe Thr Asn Val Ala Val
725 730 735
Asn Val Gly Leu Thr Val Leu Ala Ala Pro Thr Gly Thr Thr Thr Val
740 745 750
Thr Leu Ala Asn Ala Thr Gly Thr Ser Asp Val Phe Asn Leu Thr Leu
755 760 765
Ser Ser Ser Ala Ala Leu Ala Ala Gly Thr Val Ala Leu Ala Gly Val
770 775 780
Glu Thr Val Asn Ile Ala Ala Thr Asp Thr Asn Thr Thr Ala His Val
785 790 795 800
Asp Thr Leu Thr Leu Gln Ala Thr Ser Ala Lys Ser Ile Val Val Thr
805 810 815
Gly Asn Ala Gly Leu Asn Leu Thr Asn Thr Gly Asn Thr Ala Val Thr
820 825 830
Ser Phe Asp Ala Ser Ala Val Thr Gly Thr Ala Pro Ala Val Thr Phe
835 840 845
Val Ser Ala Asn Thr Thr Val Gly Glu Val Val Thr Ile Arg Gly Gly
850 855 860
Ala Gly Ala Asp Ser Leu Thr Gly Ser Ala Thr Ala Asn Asp Thr Ile
865 870 875 880
Ile Gly Gly Ala Gly Ala Asp Thr Leu Val Tyr Thr Gly Gly Thr Asp
885 890 895
Thr Phe Thr Gly Gly Thr Gly Ala Asp Ile Phe Asp Ile Asn Ala Ile
900 905 910
Gly Thr Ser Thr Ala Phe Val Thr Ile Thr Asp Ala Ala Val Gly Asp
915 920 925
Lys Leu Asp Leu Val Gly Ile Ser Thr Asn Gly Ala Ile Ala Asp Gly

930 935 940
Ala Phe Gly Ala Ala Val Thr Leu Gly Ala Ala Ala Thr Leu Ala Gln
945 950 955 960
Tyr Leu Asp Ala Ala Ala Ala Gly Asp Gly Ser Gly Thr Ser Val Ala
965 970 975
Lys Trp Phe Gln Phe Gly Gly Asp Thr Tyr Val Val Val Asp Ser Ser
980 985 990
Ala Gly Ala Thr Phe Val Ser Gly Ala Asp Ala Val Ile Lys Leu Thr
995 1000 1005
Gly Leu Val Thr Leu Thr Thr Ser Ala Phe Ala Thr Glu Val Leu Thr
1010 1015 1020
Leu Ala
1025

<210> 6
<211> 306
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetically generated polynucleotide

<221> CDS
<222> (1)...(306)

<400> 6
gaa ttc aga tct cag ggc gcg ggg cag ggt ggc tat ggt ggg ctc ggc 48
Glu Phe Arg Ser Gln Gly Ala Gly Gln Gly Gly Tyr Gly Gly Leu Gly
1 5 10 15

tcg caa ggc gct ggc ctg ggt ggc cag ggc gct ggc gcg gcc gcg gcc 96
Ser Gln Gly Ala Gly Leu Gly Gly Gln Gly Ala Gly Ala Ala Ala Ala
20 25 30

gct gcg gcc ggt ggc gct ggc cag ggc ggg ctg ggc tcg cag ggc gcc 144
Ala Ala Ala Gly Gly Ala Gly Gln Gly Gly Leu Gly Ser Gln Gly Ala
35 40 45

ggc caa ggc gct ggc gcc gcg gcc gct gcg gcc ggt ggc gcc ggc cag 192
Gly Gln Gly Ala Gly Ala Ala Ala Ala Ala Ala Gly Gly Ala Gly Gln
50 55 60

ggt ggc tac ggc ggc ctg ggc agc cag ggc gcc ggt cgc ggc ggt cag 240
Gly Gly Tyr Gly Gly Leu Gly Ser Gln Gly Ala Gly Arg Gly Gly Gln
65 70 75 80

ggc gcc ggt gcc gcg gcc gct gcg gcc ggt ggc gct ggg caa ggc ggc 288
Gly Ala Gly Ala Ala Ala Ala Ala Ala Gly Gly Ala Gly Gln Gly Gly
85 90 95

tac ggc ggt ctg gga tcc 306
Tyr Gly Gly Leu Gly Ser
100

<210> 7
<211> 102

<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetically generated polypeptide

<400> 7
Glu Phe Arg Ser Gln Gly Ala Gly Gln Gly Gly Tyr Gly Gly Leu Gly
1 5 10 15
Ser Gln Gly Ala Gly Leu Gly Gly Gln Gly Ala Gly Ala Ala Ala Ala
20 25 30
Ala Ala Ala Gly Gly Ala Gly Gln Gly Gly Leu Gly Ser Gln Gly Ala
35 40 45
Gly Gln Gly Ala Gly Ala Ala Ala Ala Ala Ala Gly Gly Ala Gly Gln
50 55 60
Gly Gly Tyr Gly Gly Leu Gly Ser Gln Gly Ala Gly Arg Gly Gly Gln
65 70 75 80
Gly Ala Gly Ala Ala Ala Ala Ala Ala Gly Gly Ala Gly Gln Gly Gly
85 90 95
Tyr Gly Gly Leu Gly Ser
100

<210> 8
<211> 780
<212> DNA
<213> Infectious Pancreatic Necrosis Virus

<220>
<221> CDS
<222> (1)...(780)

<400> 8
atg aac aca aac aag gca acc gca act tac ttg aaa tcc att atg ctt 48
Met Asn Thr Asn Lys Ala Thr Ala Thr Tyr Leu Lys Ser Ile Met Leu
1 5 10 15

cca gag act gga cca gca agc atc ccg gac gac ata acg gag aga cac 96
Pro Glu Thr Gly Pro Ala Ser Ile Pro Asp Asp Ile Thr Glu Arg His
20 25 30

atc tta aaa caa gag acc tcg tca tac aac tta gag gtc tcc gaa tca 144
Ile Leu Lys Gln Glu Thr Ser Ser Tyr Asn Leu Glu Val Ser Glu Ser
35 40 45

gga agt ggc att ctt gtt tgt ttc cct ggg gca cca ggc tca cgg atc 192
Gly Ser Gly Ile Leu Val Cys Phe Pro Gly Ala Pro Gly Ser Arg Ile
50 55 60

ggt gca cac tac aga tgg aat gcg aac cag acg ggg ctg gag ttc gac 240
Gly Ala His Tyr Arg Trp Asn Ala Asn Gln Thr Gly Leu Glu Phe Asp
65 70 75 80

cag tgg ctg gag acg tcg cag gac ctg aag aaa gcc ttc aac tac ggg 288
Gln Trp Leu Glu Thr Ser Gln Asp Leu Lys Lys Ala Phe Asn Tyr Gly
85 90 95

agg ctg atc tca agg aaa tac gac att caa agc tcc aca cta ccg gcc 336

```
Arg Leu Ile Ser Arg Lys Tyr Asp Ile Gln Ser Ser Thr Leu Pro Ala
            100                 105                 110

ggt ctc tat gct ctg aac ggg acg ctc aac gct gcc acc ttc gaa ggc       384
Gly Leu Tyr Ala Leu Asn Gly Thr Leu Asn Ala Ala Thr Phe Glu Gly
        115                 120                 125

agt ctg tct gag gtg gag agc ctg acc tac aat agc ctg atg tcc cta       432
Ser Leu Ser Glu Val Glu Ser Leu Thr Tyr Asn Ser Leu Met Ser Leu
    130                 135                 140

act acg aac ccc cag gac aaa gcc aac aac cag ctg gtg acc aaa gga       480
Thr Thr Asn Pro Gln Asp Lys Ala Asn Asn Gln Leu Val Thr Lys Gly
145                 150                 155                 160

gtc acc gtc ctg aat cta cca aca ggg ttc gac aaa cca tac gtc cgc       528
Val Thr Val Leu Asn Leu Pro Thr Gly Phe Asp Lys Pro Tyr Val Arg
                165                 170                 175

cta gag gac gag aca ccc cag ggt ctc cag tca atg aac ggg gcc agg       576
Leu Glu Asp Glu Thr Pro Gln Gly Leu Gln Ser Met Asn Gly Ala Arg
            180                 185                 190

ctg agg tgc aca gct gca att gca cca cgg agg tac gag atc gac ctc       624
Leu Arg Cys Thr Ala Ala Ile Ala Pro Arg Arg Tyr Glu Ile Asp Leu
        195                 200                 205

cca tcc caa agc cta ccc ccc gtt cct gcg aca gga acc ctc acc act       672
Pro Ser Gln Ser Leu Pro Pro Val Pro Ala Thr Gly Thr Leu Thr Thr
    210                 215                 220

ctc tac gag gga aac gcc gac atc gtc agc tcc aca aca gtg acg gga       720
Leu Tyr Glu Gly Asn Ala Asp Ile Val Ser Ser Thr Thr Val Thr Gly
225                 230                 235                 240

gac ata aac ttc agt ctg gca gaa cga ccc gca aac gag acc agg ttc       768
Asp Ile Asn Phe Ser Leu Ala Glu Arg Pro Ala Asn Glu Thr Arg Phe
                245                 250                 255

gac ttc cag ctg                                                       780
Asp Phe Gln Leu
            260


<210> 9
<211> 260
<212> PRT
<213> Infectious Pancreatic Necrosis Virus

<400> 9
Met Asn Thr Asn Lys Ala Thr Ala Thr Tyr Leu Lys Ser Ile Met Leu
 1               5                  10                  15
Pro Glu Thr Gly Pro Ala Ser Ile Pro Asp Asp Ile Thr Glu Arg His
            20                  25                  30
Ile Leu Lys Gln Glu Thr Ser Ser Tyr Asn Leu Glu Val Ser Glu Ser
        35                  40                  45
Gly Ser Gly Ile Leu Val Cys Phe Pro Gly Ala Pro Gly Ser Arg Ile
```

50 55 60
Gly Ala His Tyr Arg Trp Asn Ala Asn Gln Thr Gly Leu Glu Phe Asp
65 70 75 80
Gln Trp Leu Glu Thr Ser Gln Asp Leu Lys Lys Ala Phe Asn Tyr Gly
85 90 95
Arg Leu Ile Ser Arg Lys Tyr Asp Ile Gln Ser Ser Thr Leu Pro Ala
100 105 110
Gly Leu Tyr Ala Leu Asn Gly Thr Leu Asn Ala Ala Thr Phe Glu Gly
115 120 125
Ser Leu Ser Glu Val Glu Ser Leu Thr Tyr Asn Ser Leu Met Ser Leu
130 135 140
Thr Thr Asn Pro Gln Asp Lys Ala Asn Asn Gln Leu Val Thr Lys Gly
145 150 155 160
Val Thr Val Leu Asn Leu Pro Thr Gly Phe Asp Lys Pro Tyr Val Arg
165 170 175
Leu Glu Asp Glu Thr Pro Gln Gly Leu Gln Ser Met Asn Gly Ala Arg
180 185 190
Leu Arg Cys Thr Ala Ala Ile Ala Pro Arg Arg Tyr Glu Ile Asp Leu
195 200 205
Pro Ser Gln Ser Leu Pro Pro Val Pro Ala Thr Gly Thr Leu Thr Thr
210 215 220
Leu Tyr Glu Gly Asn Ala Asp Ile Val Ser Ser Thr Thr Val Thr Gly
225 230 235 240
Asp Ile Asn Phe Ser Leu Ala Glu Arg Pro Ala Asn Glu Thr Arg Phe
245 250 255
Asp Phe Gln Leu
260

<210> 10
<211> 131
<212> PRT
<213> Escherichia coli

<400> 10
Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
1 5 10 15
Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly Gln
20 25 30
Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
35 40 45
Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Ser
50 55 60
Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn Ser
65 70 75 80
Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val Ser
85 90 95
Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val Leu
100 105 110
Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Cys Asp
115 120 125
Val Ser Ala
130

<210> 11
<211> 131
<212> PRT
<213> Escherichia coli

<400> 11

Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
1 5 10 15

Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly Gln
20 25 30

Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
35 40 45

Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Ser
50 55 60

Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn Ser
65 70 75 80

Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val Ser
85 90 95

Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val Leu
100 105 110

Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Cys Asp
115 120 125

Val Ser Ala
130

Appendix 1

```
GCTATTGTCG ACGTATGACG TTTGCTCTAT AGCCATCGCT GCTCCCATGC GCGCCACTCG    60
GTCGCAGGGG GTGTGGATT TTTTTTGGGA GACAATCCTC ATGGCCTATA CGACGGCCCA   120
GTTGGTGACT GCGTACACCA ACGCCAACCT CGGCAAGGCG CCTGACGCCG CCACCACGCT   180
GACGCTCGAC GCGTACGCGA CTCAAACCCA GACGGGCGGC CTCTCGGACG CCGCTGCGCT   240
GACCAACACC CTGAAGCTGG TCAACAGCAC GACGGCTGTT GCCATCCAGA CCTACCAGTT   300
CTTCACCGGC GTTGCCCCGT CGGCCGCTGG TCTGGACTTC CTGGTCGACT CGACCACCAA   360
CACCAACGAC CTGAACGACG CGTACTACTC GAAGTTCGCT CAGGAAAACC GCTTCATCAA   420
CTTCTCGATC AACCTGGCCA CGGGCGCCGG CGCCGGCGCG ACGGCTTTCG CCGCCGCCTA   480
CACGGGCGTT TCGTACGCCC AGACGGTCGC CACCGCCTAT GACAAGATCA TCGGCAACGC   540
CGTCGCGACC GCCGCTGGCG TCGACGTCGC GGCCGCCGTG GCTTTCCTGA GCCGCCAGGC   600
CAACATCGAC TACCTGACCG CCTTCGTGCG CGCCAACACG CCGTTCACGG CCGCTGCCGA   660
CATCGATCTG GCCGTCAAGG CCGCCCTGAT CGGCACCATC CTGAACGCCG CCACGGTGTC   720
GGGCATCGGT GGTTACGCGA CCGCCACGGC CGCGATGATC AACGACCTGT CGGACGGCGC   780
CCTGTCGACC GACAACGCGG CTGGCGTGAA CCTGTTCACC GCCTATCCGT CGTCGGGCGT   840
GTCGGGTTCG ACCCTCTCGC TGACCACCGG CACCGACACC CTGACGGGCA CCGCCAACAA   900
CGACACGTTC GTTGCGGGTG AAGTCGCCGG CGCTGCGACC CTGACCGTTG GCGACACCCT   960
GAGCGGCGGT GCTGGCACCG ACGTCCTGAA CTGGGTGCAA GCTGCTGCGG TTACGGCTCT  1020
GCCGACCGGC GTGACGATCT CGGGCATCGA AACGATGAAC GTGACGTCGG GCGCTGCGAT  1080
CACCCTGAAC ACGTCTTCGG GCGTGACGGG TCTGACCGCC CTGAACACCA ACACCAGCGG  1140
CGCGGCTCAA ACCGTCACCG CCGGCGCTGG CCAGAACCTG ACCGCCACGA CCGCCGCTCA  1200
AGCCGCGAAC AACGTCGCCG TCGACGGGCG CGCCAACGTC ACCGTCGCCT CGACGGGCGT  1260
GACCTCGGGC ACGACCACGG TCGGCGCCAA CTCGGCCGCT TCGGGCACCG TGTCGGTGAG  1320
CGTCGCGAAC TCGAGCACGA CCACCACGGG CGCTATCGCC GTGACCGGTG GTACGGCCGT  1380
GACCGTGGCT CAAACGGCCG GCAACGCCGT GAACACCACG TTGACGCAAG CCGACGTGAC  1440
CGTGACCGGT AACTCCAGCA CCACGGCCGT GACGGTCACC CAAACCGCCG CCGCCACCGC  1500
CGGCGCTACG GTCGCCGGTC GCGTCAACGG CGCTGTGACG ATCACCGACT CTGCCGCCGC  1560
CTCGGCCACG ACCGCCGGCA AGATCGCCAC GGTCACCCTG GGCAGCTTCG GCGCCGCCAC  1620
GATCGACTCG AGCGCTCTGA CGACCGTCAA CCTGTCGGGC ACGGGCACCT CGCTCGGCAT  1680
```

Appendix 1 (cont'd)

CGGCCGCGGC GCTCTGACCG CCACGCCGAC CGCCAACACC CTGACCCTGA ACGTCAATGG 1740
TCTGACGACG ACCGGCGCGA TCACGGACTC GGAAGCGGCT GCTGACGATG GTTTCACCAC 1800
CATCAACATC GCTGGTTCGA CCGCCTCTTC GACGATCGCC AGCCTGGTGG CCGCCGACGC 1860
GACGACCCTG AACATCTCGG GCGACGCTCG CGTCACGATC ACCTCGCACA CCGCTGCCGC 1920
CCTGACGGGC ATCACGGTGA CCAACAGCGT TGGTGCGACC CTCGGCGCCG AACTGGCGAC 1980
CGGTCTGGTC TTCACGGGCG GCGCTGGCCG TGACTCGATC CTGCTGGGCG CCACGACCAA 2040
GGCGATCGTC ATGGGCGCCG GCGACGACAC CGTCACCGTC AGCTCGGCGA CCCTGGGCGC 2100
TGGTGGTTCG GTCAACGGCG GCGACGGCAC CGACGTTCTG GTGGCCAACG TCAACGGTTC 2160
GTCGTTCAGC GCTGACCCGG CCTTCGGCGG CTTCGAAACC CTCCGCGTCG CTGGCGCGGC 2220
GGCTCAAGGC TCGCACAACG CCAACGGCTT CACGGCTCTG CAACTGGGCG CGACGGCGGG 2280
TGCGACGACC TTCACCAACG TTGCGGTGAA TGTCGGCCTG ACCGTTCTGG CGGCTCCGAC 2340
CGGTACGACG ACCGTGACCC TGGCCAACGC CACGGGCACC TCGGACGTGT TCAACCTGAC 2400
CCTGTCGTCC TCGGCCGCTC TGGCCGCTGG TACGGTTGCG CTGGCTGGCG TCGAGACGGT 2460
GAACATCGCC GCCACCGACA CCAACACGAC CGCTCACGTC GACACGCTGA CGCTGCAAGC 2520
CACCTCGGCC AAGTCGATCG TGGTGACGGG CAACGCCGGT CTGAACCTGA CCAACACCGG 2580
CAACACGGCT GTCACCAGCT TCGACGCCAG CGCCGTCACC GGCACGGCTC CGGCTGTGAC 2640
CTTCGTGTCG GCCAACACCA CGGTGGGTGA AGTCGTCACG ATCCGCGGCG GCGCTGGCGC 2700
CGACTCGCTG ACCGGTTCGG CCACCGCCAA TGACACCATC ATCGGTGGCG CTGGCGCTGA 2760
CACCCTGGTC TACACCGGCG GTACGGACAC CTTCACGGGT GGCACGGGCG CGGATATCTT 2820
CGATATCAAC GCTATCGGCA CCTCGACCGC TTTCGTGACG ATCACCGACG CCGCTGTCGG 2880
CGACAAGCTC GACCTCGTCG GCATCTCGAC GAACGGCGCT ATCGCTGACG GCGCCTTCGG 2940
CGCTGCGGTC ACCCTGGGCG CTGCTGCGAC CCTGGCTCAG TACCTGGACG CTGCTGCTGC 3000
CGGCGACGGC AGCGGCACCT CGGTTGCCAA GTGGTTCCAG TTCGGCGGCG ACACCTATGT 3060
CGTCGTTGAC AGCTCGGCTG GCGCGACCTT CGTCAGCGGC GCTGACGCGG TGATCAAGCT 3120
GACCGGTCTG GTCACGCTGA CCACCTCGGC CTTCGCCACC GAAGTCCTGA CGCTCGCCTA 3180
AGCGAACGTC TGATCCTCGC CTAGGCGAGG ATCGCTAGAC TAAGAGACCC CGTCTTCCGA 3240
AAGGGAGGCG GGGTCTTTCT TATGGGCGCT ACGCGCTGGC CGGCCTTGCC TAGTTCCGGT 3300

Appendix 1 (cont'd)

```
Met Ala Tyr Thr Thr Ala Gln Leu Val Thr Ala Tyr Thr Asn Ala Asn
1               5                   10                  15

Leu Gly Lys Ala Pro Asp Ala Ala Thr Thr Leu Thr Leu Asp Ala Tyr
            20                  25                  30

Ala Thr Gln Thr Gln Thr Gly Gly Leu Ser Asp Ala Ala Ala Leu Thr
        35                  40                  45

Asn Thr Leu Lys Leu Val Asn Ser Thr Thr Ala Val Ala Ile Gln Thr
    50                  55                  60

Tyr Gln Phe Phe Thr Gly Val Ala Pro Ser Ala Ala Gly Leu Asp Phe
65                  70                  75                  80

Leu Val Asp Ser Thr Thr Asn Thr Asn Asp Leu Asn Asp Ala Tyr Tyr
                85                  90                  95

Ser Lys Phe Ala Gln Glu Asn Arg Phe Ile Asn Phe Ser Ile Asn Leu
            100                 105                 110

Ala Thr Gly Ala Gly Ala Gly Ala Thr Ala Phe Ala Ala Ala Tyr Thr
        115                 120                 125

Gly Val Ser Tyr Ala Gln Thr Val Ala Thr Ala Tyr Asp Lys Ile Ile
    130                 135                 140

Gly Asn Ala Val Ala Thr Ala Ala Gly Val Asp Val Ala Ala Ala Val
145                 150                 155                 160

Ala Phe Leu Ser Arg Gln Ala Asn Ile Asp Tyr Leu Thr Ala Phe Val
                165                 170                 175

Arg Ala Asn Thr Pro Phe Thr Ala Ala Ala Asp Ile Asp Leu Ala Val
            180                 185                 190

Lys Ala Ala Leu Ile Gly Thr Ile Leu Asn Ala Ala Thr Val Ser Gly
        195                 200                 205

Ile Gly Gly Tyr Ala Thr Ala Thr Ala Ala Met Ile Asn Asp Leu Ser
    210                 215                 220

Asp Gly Ala Leu Ser Thr Asp Asn Ala Ala Gly Val Asn Leu Phe Thr
225                 230                 235                 240

Ala Tyr Pro Ser Ser Gly Val Ser Gly Ser Thr Leu Ser Leu Thr Thr
                245                 250                 255

Gly Thr Asp Thr Leu Thr Gly Thr Ala Asn Asn Asp Thr Phe Val Ala
            260                 265                 270

Gly Glu Val Ala Gly Ala Ala Thr Leu Thr Val Gly Asp Thr Leu Ser
        275                 280                 285

Gly Gly Ala Gly Thr Asp Val Leu Asn Trp Val Gln Ala Ala Ala Val
    290                 295                 300

Thr Ala Leu Pro Thr Gly Val Thr Ile Ser Gly Ile Glu Thr Met Asn
305                 310                 315                 320

Val Thr Ser Gly Ala Ala Ile Thr Leu Asn Thr Ser Ser Gly Val Thr
                325                 330                 335

Gly Leu Thr Ala Leu Asn Thr Asn Thr Ser Gly Ala Ala Gln Thr Val
            340                 345                 350
```

Appendix 1 (cont'd)

```
Thr Ala Gly Ala Gly Gln Asn Leu Thr Ala Thr Thr Ala Ala Gln Ala
        355                 360                 365

Ala Asn Asn Val Ala Val Asp Gly Arg Ala Asn Val Thr Val Ala Ser
    370                 375                 380

Thr Gly Val Thr Ser Gly Thr Thr Thr Val Gly Ala Asn Ser Ala Ala
385                 390                 395                 400

Ser Gly Thr Val Ser Val Ser Val Ala Asn Ser Ser Thr Thr Thr Thr
                405                 410                 415

Gly Ala Ile Ala Val Thr Gly Gly Thr Ala Val Thr Val Ala Gln Thr
            420                 425                 430

Ala Gly Asn Ala Val Asn Thr Thr Leu Thr Gln Ala Asp Val Thr Val
        435                 440                 445

Thr Gly Asn Ser Ser Thr Thr Ala Val Thr Val Thr Gln Thr Ala Ala
    450                 455                 460

Ala Thr Ala Gly Ala Thr Val Ala Gly Arg Val Asn Gly Ala Val Thr
465                 470                 475                 480

Ile Thr Asp Ser Ala Ala Ala Ser Ala Thr Thr Ala Gly Lys Ile Ala
                485                 490                 495

Thr Val Thr Leu Gly Ser Phe Gly Ala Ala Thr Ile Asp Ser Ser Ala
            500                 505                 510

Leu Thr Thr Val Asn Leu Ser Gly Thr Gly Thr Ser Leu Gly Ile Gly
        515                 520                 525

Arg Gly Ala Leu Thr Ala Thr Pro Thr Ala Asn Thr Leu Thr Leu Asn
    530                 535                 540

Val Asn Gly Leu Thr Thr Thr Gly Ala Ile Thr Asp Ser Glu Ala Ala
545                 550                 555                 560

Ala Asp Asp Gly Phe Thr Thr Ile Asn Ile Ala Gly Ser Thr Ala Ser
                565                 570                 575

Ser Thr Ile Ala Ser Leu Val Ala Ala Asp Ala Thr Thr Leu Asn Ile
            580                 585                 590

Ser Gly Asp Ala Arg Val Thr Ile Thr Ser His Thr Ala Ala Ala Leu
        595                 600                 605

Thr Gly Ile Thr Val Thr Asn Ser Val Gly Ala Thr Leu Gly Ala Glu
    610                 615                 620

Leu Ala Thr Gly Leu Val Phe Thr Gly Gly Ala Gly Arg Asp Ser Ile
625                 630                 635                 640

Leu Leu Gly Ala Thr Thr Lys Ala Ile Val Met Gly Ala Gly Asp Asp
                645                 650                 655

Thr Val Thr Val Ser Ser Ala Thr Leu Gly Ala Gly Gly Ser Val Asn
            660                 665                 670

Gly Gly Asp Gly Thr Asp Val Leu Val Ala Asn Val Asn Gly Ser Ser
        675                 680                 685

Phe Ser Ala Asp Pro Ala Phe Gly Gly Phe Glu Thr Leu Arg Val Ala
    690                 695                 700
```

Appendix 1 (cont'd)

Gly Ala Ala Ala Gln Gly Ser His Asn Ala Asn Gly Phe Thr Ala Leu
705 710 715 720

Gln Leu Gly Ala Thr Ala Gly Ala Thr Thr Phe Thr Asn Val Ala Val
725 730 735

Asn Val Gly Leu Thr Val Leu Ala Ala Pro Thr Gly Thr Thr Thr Val
740 745 750

Thr Leu Ala Asn Ala Thr Gly Thr Ser Asp Val Phe Asn Leu Thr Leu
755 760 765

Ser Ser Ser Ala Ala Leu Ala Ala Gly Thr Val Ala Leu Ala Gly Val
770 775 780

Glu Thr Val Asn Ile Ala Ala Thr Asp Thr Asn Thr Thr Ala His Val
785 790 795 800

Asp Thr Leu Thr Leu Gln Ala Thr Ser Ala Lys Ser Ile Val Val Thr
805 810 815

Gly Asn Ala Gly Leu Asn Leu Thr Asn Thr Gly Asn Thr Ala Val Thr
820 825 830

Ser Phe Asp Ala Ser Ala Val Thr Gly Thr Ala Pro Ala Val Thr Phe
835 840 845

Val Ser Ala Asn Thr Thr Val Gly Glu Val Val Thr Ile Arg Gly Gly
850 855 860

Ala Gly Ala Asp Ser Leu Thr Gly Ser Ala Thr Ala Asn Asp Thr Ile
865 870 875 880

Ile Gly Gly Ala Gly Ala Asp Thr Leu Val Tyr Thr Gly Gly Thr Asp
885 890 895

Thr Phe Thr Gly Gly Thr Gly Ala Asp Ile Phe Asp Ile Asn Ala Ile
900 905 910

Gly Thr Ser Thr Ala Phe Val Thr Ile Thr Asp Ala Ala Val Gly Asp
915 920 925

Lys Leu Asp Leu Val Gly Ile Ser Thr Asn Gly Ala Ile Ala Asp Gly
930 935 940

Ala Phe Gly Ala Ala Val Thr Leu Gly Ala Ala Ala Thr Leu Ala Gln
945 950 955 960

Tyr Leu Asp Ala Ala Ala Ala Gly Asp Gly Ser Gly Thr Ser Val Ala
965 970 975

Lys Trp Phe Gln Phe Gly Gly Asp Thr Tyr Val Val Val Asp Ser Ser
980 985 990

Ala Gly Ala Thr Phe Val Ser Gly Ala Asp Ala Val Ile Lys Leu Thr
995 1000 1005

Gly Leu Val Thr Leu Thr Thr Ser Ala Phe Ala Thr Glu Val Leu Thr
1010 1015 1020

Leu Ala
1025

Appendix 2

GAA TTC AGA TCT CAG GGC GCG GGG CAG GGT GGC TAT GGT GGG CTC GGC
TCG CAA GGC
GCT
E F R S Q G A G Q G G Y G G L G S Q G A

GGC CTG GGT GGC CAG GGC GCT GGC GCG GCC GCG GCC GCT GCG GCC GGT
GGC
G R G G Q G A G A A A A A A A G G

GCT GGC CAG GGC GGG CTG GGC TCG CAG GGC GCC GGC CAA GGC GCT GGC
GCC GCG GCC
GCT
A G Q G G L G S Q G A G Q G A G A A A A

GCG GCC GGT GGC GCC GGC CAG GGT GGC TAC GGC GGC CTG GGC AGC CAG
GGC GCC GGT
CGC
A A G G A G Q G G Y G G L G S Q G A G R

GGC GGT CAG GGC GCC GGT GCC GCG GCC GCT GCG GCC GGT GGC GCT GGG
CAA GGC GGC TAC
G G Q G A G A A A A A A G G A G Q G G Y

GGC GGT CTG GGA TCC
G G L G S

1/1

# Appendix 3

atg aac aca aac aag gca acc gca act tac ttg aaa tcc att atg ctt cca gag act
gga
Met asn thr asn lys ala thr ala thr tyr leu lys ser ile met leu pro glu thr
gly
61/21

cca gca agc atc ccg gac gac ata acg gag aga cac atc tta aaa caa gag acc tcg
tca
pro ala ser ile pro asp asp ile thr glu arg his ile leu lys gln glu thr ser
ser
121/41

tac aac tta gag gtc tcc gaa tca gga agt ggc att ctt gtt tgt ttc cct ggg gca
cca
tyr asn leu glu val ser glu ser gly ser gly ile leu val cys phe pro gly ala
pro
181/61

ggc tca cgg atc ggt gca cac tac aga tgg aat grg aac cag acg ggg ctg gag ttc
gac
gly ser arg ile gly ala his tyr arg trp asn ala asn gln thr gly leu glu phe
asp
241/81

cag tgg ctg gag acg tcg cag gac ctg aag aaa gcc ttc aac tac ggg agg ctg atc
tca
gln trp leu glu thr ser gln asp leu lys lys ala phe asn tyr gly arg leu ile
ser
301/101

agg aaa tac gac att caa agc tcc aca cta ccg gcc ggt ctc tat gct ctg aac ggc
acg
arg lys tyr asp ile gln ser ser thr leu pro ala gly leu tyr ala leu asn gly
thr
361/121

ctc aac gct gcc acc ttc gaa ggc agt ctg tct gag gtg gag agc ctg acc tac aat
agc
leu asn ala ala thr phe glu gly ser leu ser glu val glu ser leu thr tyr asn
ser
421/141

ctg atg tcc cta act acg aac ccc cag gac aaa gcc aac aac cag ctg gtg acc aaa
gga
leu met ser leu thr thr asn pro gln asp lys ala asn asn gln leu val thr lys
gly
481/161

gtc acc gtc ctg aat cta cca aca ggg ttc gac aaa cca tac gtc cgc cta gag gac
gag
val thr val leu asn leu pro thr gly phe asp lys pro tyr val arg leu glu asp
glu
541/181

aca ccc cag ggt ctc cag tca atg aac ggg gcc agg atg agg tgc aca gct gca att
gca
thr pro gln gly leu gln ser met asn gly ala arg met arg cys thr ala ala ile
ala
601/201

cca cgg agg tac gag atc gac ctc cca tcc caa agc cta ccc ccc gtt cct gcg aca
gga
pro arg arg tyr glu ile asp leu pro ser gln ser leu pro pro val pro ala thr
gly
661/221

acc ctc acc act ctc tac gag gga aac gcc gac atc gtc agc tcc aca aca gtg acg
gga
thr leu thr thr leu tyr glu gly asn ala asp ile val ser ser thr thr val thr
gly
721/241

gac ata aac ttc agt ctg gca gaa cga ccc gca aac gag acc agg ttc gac ttc cag
ctg
asp ile asn phe ser leu ala glu arg pro ala asn glu thr arg phe asp phe gln
leu

Appendix 4

The T3 protein sequence is:
FACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSA
SYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQ
CDVSA

The T7 protein sequence is:
FACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSA
SYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQ
CDVSARDVTVTLPDYRGSVPIPLTVYCAKSQNLGYYLSGTHADAGNSIFTNTASFSPAQGVG
GAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQ

**WHAT IS CLAIMED IS:**

1. A method of cleaving a fusion protein including a first component which comprises all or part of a Caulobacter S-layer protein including a Caulobacter C-terminal secretion signal, and a second component heterologous to Caulobacter, the fusion protein containing at least one aspartate-proline dipeptide, wherein the method comprises combining the fusion protein with an acid solution of a strength insufficient to solubilize the fusion protein for a time sufficient for cleavage of the fusion protein at said aspartate-proline dipeptide.

2. The method of claim 1 wherein a aspartate-proline dipeptide is situated between the first and second components or adjacent a junction between the first and second components.

3. The method of claim 1 or 2, wherein the acid solution has a pH of from about 1.5 to about 2.5.

4. The method of claim 1 or 2, wherein the acid solution has a pH of about 1.65 to about 2.35.

5. The method of any one of claims 1-4 wherein the method is carried out at a temperature in the range of about 30° C. to about 50° C.

6. The method of any one of claims 1-5, wherein the method further comprises separating products cleaved from the fusion protein.

7. A method of preparing a DNA construct for expression of a fusion protein suitable for use in the method of claim 1, wherein the method comprises joining an upstream DNA segment including DNA heterologous to Caulobacter which encodes a protein

of interest, to a downstream DNA segment including DNA for a Caulobacter C-terminal secretion signal, wherein the downstream DNA segment does not encode an aspartate-proline dipeptide, and wherein the upstream segment contains DNA encoding an aspartate-proline dipeptide at or near an end of said upstream segment to be joined to said downstream segment.

8. A method of preparing a fusion protein, comprising:

(1) expressing a DNA construct prepared as described in claim 7 in Caulobacter and,

(2) recovering said fusion protein secreted by the Caulobacter.

# COMBINED DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled CLEAVAGE OF CAULOBACTER PRODUCED RECOMBINANT FUSION PROTEINS, the specification of which:

[] is attached hereto.
[X] was filed on January 12, 2001 as Application Serial No. __________ and was amended on __________.
[X] was described and claimed in PCT International Application No. PCT/CA99/00637 filed on July 14, 1999 and as amended under PCT Article 19 on __________.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose all information I know to be material to patentability in accordance with Title 37, Code of Federal Regulations, §1.56. I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose all information I know to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56(a) which became available between the filing date of the prior application and the national or PCT international filing date of this application:

| U.S. Serial No. | Filing Date | Status |
|---|---|---|
| PCT/CA99/00637 | 14 July 1999 | Published |

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate filed by me on the same subject matter having a filing date before that of the application(s) of which priority is claimed:

| Country | Application No. | Filing Date | Priority Claimed |
|---|---|---|---|
| Canada | 2,237,704 | July 14, 1998 | [X] Yes [] No |

I hereby appoint the following attorneys and/or agents to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith: Y. Rocky Tsao, Reg. 34,053; Eldora L. Ellison, Reg. 39,967; David E. Johnson, Reg. 41,874; John T. Li, Reg. 44,210; Frank R. Occhiuti, Reg. 35,306; Eric L. Prahl, Reg. 32,590; and Gary A. Walpert, Reg. 26,098.

Address all telephone calls to Y. ROCKY TSAO at telephone number (617) 542-5070.

Address all correspondence to Y. ROCKY TSAO at:

FISH & RICHARDSON P.C.
225 Franklin Street
Boston, MA 02110-2804

## Combined Declaration and Power of Attorney



I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patents issued thereon.

Full Name of Inventor: JOHN SMIT

Inventor's Signature: ______________________ Date: 8 Mar 2001

Residence Address: 9960 Seacastle Drive
Richmond, British Columbia BCC
V7A 4R8

Citizenship: United States

Post Office Address: 9960 Seacastle Drive
Richmond, British Columbia
V7A 4R8

20165902.doc